SN Computer Science (2020) 1:85
https://doi.org/10.1007/s42979-020-0101-1

ORIGINAL RESEARCH

Cerebral LSTM: A Better Alternative for Single- and Multi-Stacked LSTM Cell-Based RNNs

Ravin Kumar[1]

© Springer Nature Singapore Pte Ltd 2020

Abstract
Deep learning has rapidly transformed the natural language processing domain with its recurrent neural networks. LSTM is one such popular repeating cell unit used for building these recurrent neural network-based deep learning architectures. In this paper, we proposed a significantly improved version of LSTM named Cerebral LSTM which has much better ability to understand time-series data. Extensive experiments were conducted to get an unbiased performance comparison of our proposed version. Obtained results showed that recurrent neural network constructed using single Cerebral LSTM cell outperformed both recurrent neural network with single LSTM cell and recurrent neural network with two-stacked LSTM cells.

Keywords Recurrent neural network · Long short-term memory · Stacked LSTM · Natural language processing

Introduction

Long short-term memory [1] has accelerated the research work for problems based on time-series data by providing solution to vanishing and exploding gradient problems of recurrent neural networks [2]. LSTM is a special type of block which requires cell state $c(t-1)$ and hidden state $h(t-1)$ along with input data $i(t)$ at each timestamp 't' to perform its operations. Fundamentally, LSTM consists of three type of gates, namely forget gate $f(t)$, input gate $i(t)$ and output gate $o(t)$ which decides relevant and irrelevant information from the input data (Fig. 1).

$$f(t) = \sigma(Wf \cdot [h(t-1), x(t)] + bf)$$
$$i(t) = \sigma(Wi \cdot [h(t-1), x(t)] + bi)$$
$$Ctmp(t) = \tanh(Wc \cdot [h(t-1), x(t)] + bc)$$
$$C(t) = f(t) * C(t-1) + i(t) * Ctmp(t)$$
$$o(t) = \sigma(Wo \cdot [h(t-1), x(t)] + bo)$$
$$h(t) = o(t) * \tanh(C(t))$$

Forget gate decides which previous information $c(t-1)$ is not required at the moment, input gate selects relevant information from the input data $x(t)$, and output gate produces new the hidden state $h(t)$ for time 't.' At each timestamp 't,' $h(t)$ also serves as the output produced by the long short-term cell for timestamp 't.'

In this paper, we have proposed a new type of recurrent cell 'Cerebral LSTM.' To show effectiveness of out proposed cell, we have conducted experiments to perform its comparative analysis with LSTM-based recurrent neural networks.

Related Works

Hochreiter et al. [1] proposed a solution for understanding long-term dependencies in recurrent neural network. Chung et al. [3] designed a recurrent unit named GRU, having performance similar to LSTM. Bidirectional LSTM developed by Graves et al. [4] showed better performance in understanding time-series data than unidirectional LSTM. Cheng et al. [5] utilized a mechanism proposed by Srivastava et al. [6] in his research work to optimize the performance of LSTM. LSTM-based recurrent neural networks were used in designing many end-to-end deep learning solutions. Huang et al. [7] used two layers LSTM-based generative model for music generation. In speech recognition, Graves et al. [8] used LSTM-based recurrent neural network for better

This article is part of the topical collection "Advances in Computational Intelligence, Paradigms and Applications" guest edited by Young Lee and S. Meenakshi Sundaram.

✉ Ravin Kumar
ravin.kumar.cs.2013@miet.ac.in

[1] Department of Computer Science, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh 250005, India

Published online: 14 March 2020

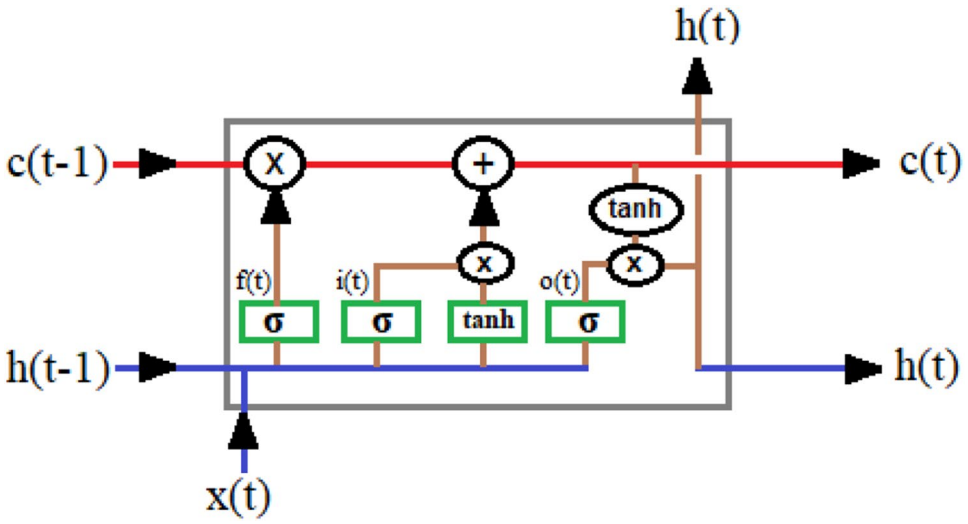

Fig. 1 Architecture of long short-term memory (LSTM) cell

performance on TIMIT phoneme recognition. Sutskever et al. [9] used LSTM cells as a basic unit in recurrent neural network of both encoder and decoder parts of sequence-to-sequence model for performing language translation. Even in other cross-domain task such as image captioning [10], LSTMs are used in decoder part for generating textual description of the input image.

Recurrent Neural Networks

In the field of deep learning, neural networks had helped in solving many problems, but they were unable to analyze the time-series data. This problem leads to the development of a new type of neural network family called 'recurrent neural networks' (Fig. 2). With further research in the field, LSTM and then later GRU-based recurrent cells were introduced to solve the vanishing and exploding gradient problem of a simple recurrent neural network (Fig. 3).

Many end-to-end deep learning architectures were developed using recurrent neural networks to efficiently solve the problems related to time- series data. For better analysis of large time-series data, mechanisms of stacking (Fig. 4), and bidirectional RNN cells (Fig. 5) were developed.

Specially, the development of sequence-to-sequence model [9] provided a huge boost in the field of natural language processing by providing end-to-end deep learning solution of various problem statements including language translation and designing conversational agents (Fig. 6).

Cerebral LSTM

Our proposed recurrent unit cell consists of one hidden state $h(t)$ and two cell states $UC(t)$ and $LC(t)$, where for each timestamp 't' we provide input $x(t)$ with hidden state $h(t-1)$ and cell states $UC(t-1)$ and $LC(t-1)$. It is called 'Cerebral LSTM' because of the similarities present in abstract architecture of cerebral hemispheres of human brain and our proposed cell (Fig. 7, Table 1).

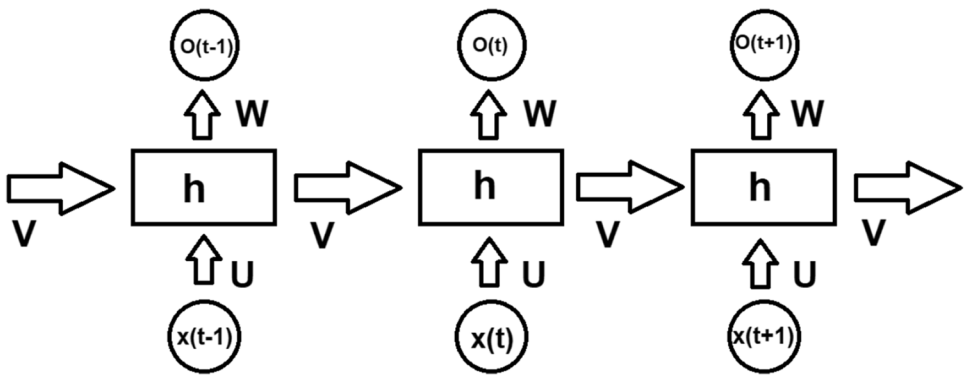

Fig. 2 Architecture of recurrent neural networks

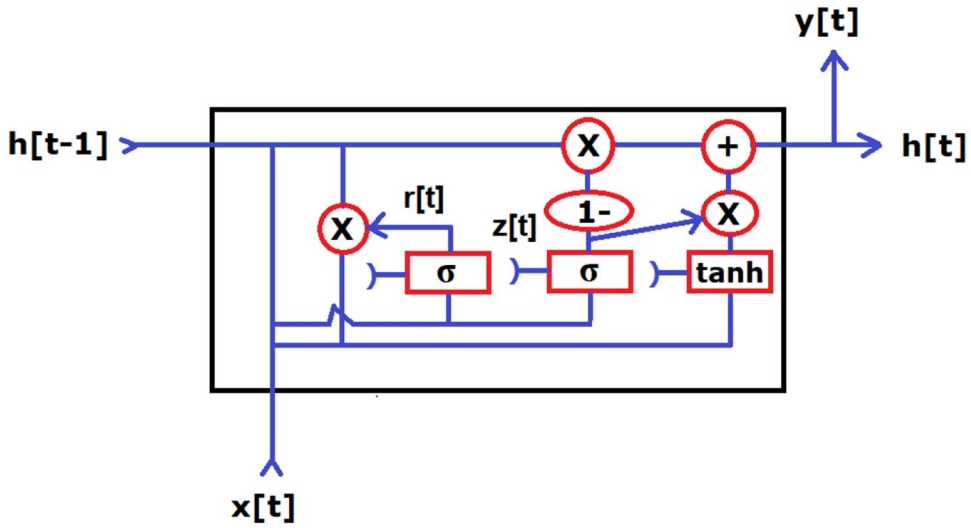

Fig. 3 Architecture of a gated recurrent unit (GRU) cell

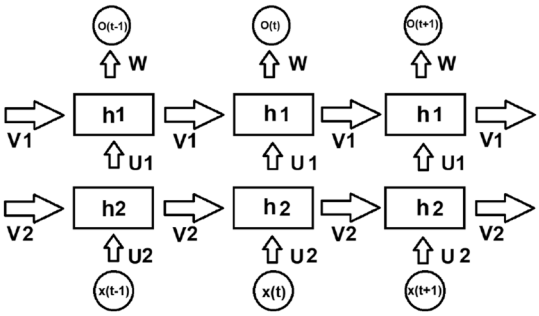

Fig. 4 Architecture of two-stacked RNN cells

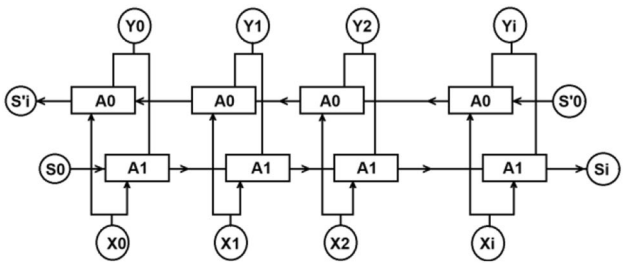

Fig. 5 Architecture of bidirectional RNNs

$$Uf(t) = \sigma(Wuf \cdot [h(t-1), x(t)] + buf)$$
$$Ui(t) = \sigma(Wui \cdot [h(t-1), x(t)] + bui)$$
$$UCtmp(t) = \tanh(Wuc \cdot [h(t-1), x(t)] + buc)$$
$$UC(t) = Uf(t) * UC(t-1) + Ui(t) * UCtmp(t)$$
$$Uo(t) = \sigma(Wuo \cdot [h(t-1), x(t)] + buo)$$
$$Lf(t) = \sigma(Wlf \cdot [h(t-1), x(t)] + blf)$$
$$Li(t) = \sigma(Wli \cdot [h(t-1), x(t)] + bli)$$
$$LCtmp(t) = \tanh(Wlc \cdot [h(t-1), x(t)] + blc)$$
$$LC(t) = Lf(t) * LC(t-1) + Li(t) * LCtmp(t)$$
$$Lo(t) = \sigma(Wlo \cdot [h(t-1), x(t)] + blo)$$
$$h(t) = Uo(t) * \tanh(UC(t)) + Lo(t) * \tanh(LC(t))$$

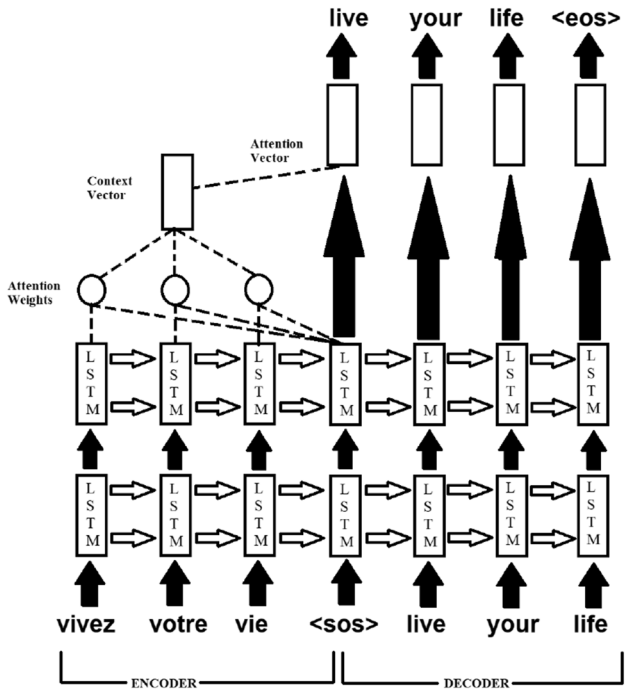

Fig. 6 Attention-based sequence-to-sequence model for language translation

In human brain, longitudinal fissure separates the cerebral hemispheres into left and right cerebral hemispheres (Fig. 8). Similarly, Cerebral LSTM consists of two cell states: UC and LC connected to same input $x(t)$ and hidden state $h(t-1)$ to update their cell states ($UC(t)$ and $LC(t)$) and jointly determine the updated value of the hidden state $h(t)$.

We have also studied the abstract brain design of various other mammals and found that even species with some level of developed intellectual abilities contains similar kind of abstract representations (Figs. 9, 10, 11).

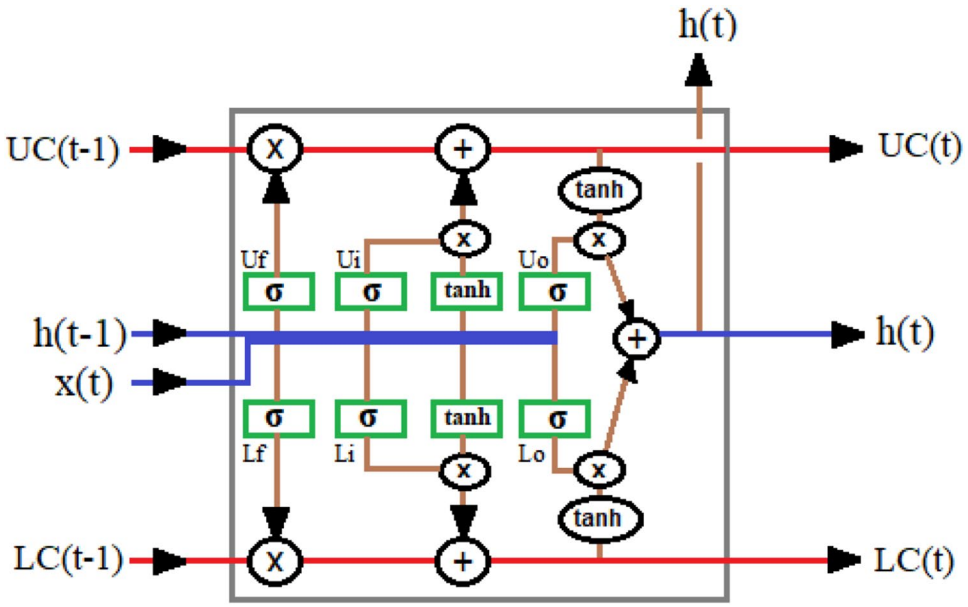

Fig. 7 Architecture of our proposed Cerebral LSTM cell

Table 1 Variables-related details

S. no.	Variables/functions	Description
1	*UC*(*t*)	Upper cell state at timestamp '*t*'
2	*LC*(*t*)	Lower cell state at timestamp '*t*'
3	*Uf*(*t*)	Upper forget gate at timestamp '*t*'
4	*Lf*(*t*)	Lower forget gate at timestamp '*t*'
5	*Ui*(*t*)	Upper input gate at timestamp '*t*'
6	*Li*(*t*)	Lower input gate at timestamp '*t*'
7	*Uo*(*t*)	Upper output gate at timestamp 's'
8	*Lo*(*t*)	Lower output gate at timestamp '*t*'
9	*h*(*t*)	Hidden state at timestamp '*t*'
10	*x*(*t*)	Input data at timestamp '*t*'

To maintain fairness in the comparative analysis among Cerebral LSTM and traditional LSTM, we have also considered the fact that our proposed recurrent cell has two cell states and an additional comparative analysis with two-stacked traditional LSTM is also performed.

Comparative Analysis

We performed a comparative study on the performance of single-LSTM and two-stacked LSTM with respect to the performance of our proposed cell using Simpson dataset [11] and then analyzed the quality of data generated by each model. To obtain unbiased results, some parameters were made constant in each comparison (Table 2).

Comparative Study of 1 LSTM with two-Stacked LSTM on the Dataset

We first studied the behavior of recurrent neural networks based on single LSTM cell and on two-stacked LSTM cells and then made comparison on the basis training loss (Fig. 12).

After 250 epochs, two-stacked LSTM-based recurrent neural network started performing better than single-LSTM-based recurrent neural network. This assures that dataset which we are using for comparative analysis is of good quality to perform further comparisons because common notion is that two-stacked LSTM should outperform single-LSTM-based recurrent neural network on dataset of considerable size.

Comparative Study of Single LSTM with Cerebral LSTM

Our proposed Cerebral LSTM showed lower training loss from the beginning as compared to recurrent neural network based on single LSTM cell which helps it better understand time-series dataset.

Up to 250 epochs, traditional LSTM showed lower training loss than two-stacked LSTM in the previous comparative study. When single LSTM cell is compared with our proposed Cerebral LSTM, it is showed that Cerebral LSTM completely outperforms single-LSTM-based recurrent neural network and maintains lower training loss from the very beginning of training phase (Fig. 13).

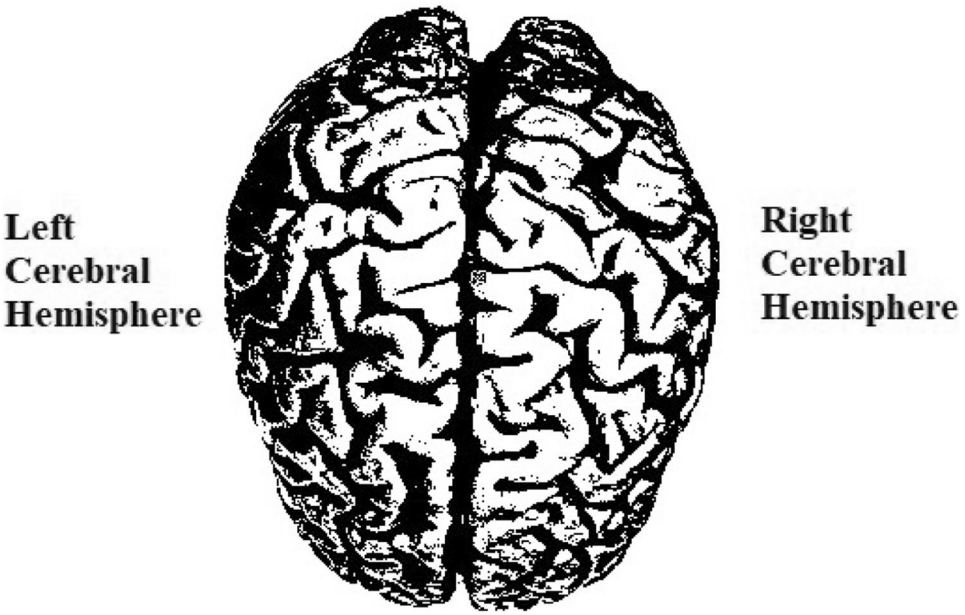

Fig. 8 Cerebral hemispheres of human brain

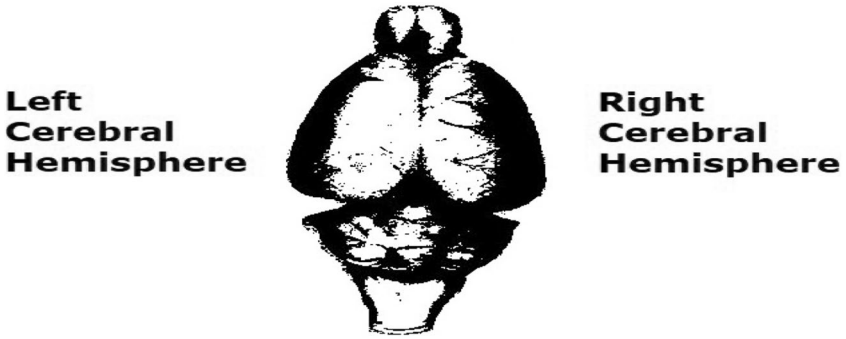

Fig. 9 Cerebral hemispheres of a rat brain

Fig. 10 Cerebral hemispheres of a sheep brain

Fig. 11 Cerebral hemispheres of a chimpanzee brain

Table 2 Parameters-related details

S. no.	Variable name	Default value
1	Learning rate	0.05
2	Epoch	500
3	Batch size	128
4	Weight initializer	Xavier initializer
5	Optimization	Adagrad optimizer
6	State size	512

Comparative Study of Two-Stacked LSTM with Cerebral LSTM

Cerebral LSTM consists of two cell states ($UC(t)$ and $LC(t)$), so we performed another comparative study to see if our proposed cell has an advantage over two-stacked LSTM-based recurrent neural networks (Fig. 14).

It can be easily seen that Cerebral LSTM easily outperformed two-stacked LSTM-based recurrent neural network. We even conducted further analysis to know whether after 500 epochs two-stacked LSTM outperforms our proposed cell, but it does not happen. The value of training loss of Cerebral LSTM on 500 epochs was 0.3979, which two-stacked LSTM achieved after 678 epochs. This analysis makes it very clear that Cerebral LSTM outperformed two-stacked LSTM.

Comparative Study of Generated Data

Generated data are also taken into consideration in our comparative study. Data generated by Cerebral LSTM were having better quality than data generated by two-stacked LSTM- and single-LSTM-based recurrent neural networks. This is because after 500 epochs Cerebral LSTM had lower training loss than two-stacked and single LSTMs which made it easier for our proposed cell to better understand the input data during training phase (Table 3).

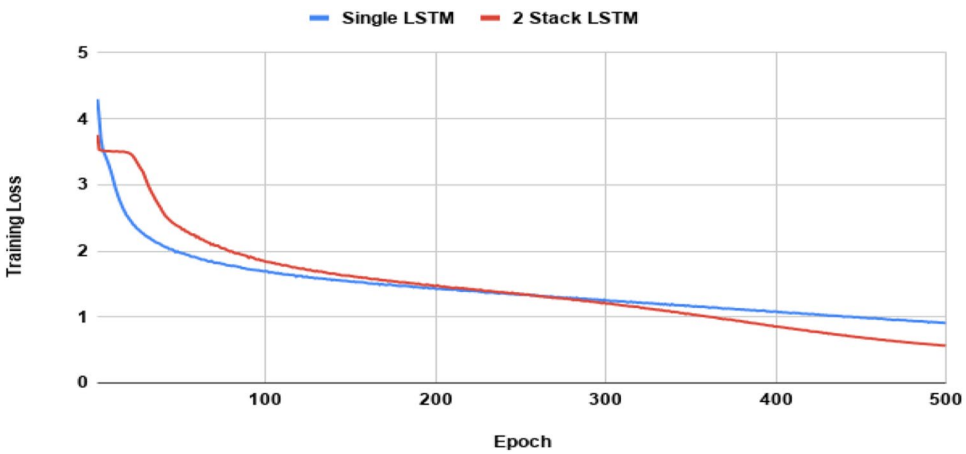

Fig. 12 Comparative study of training loss

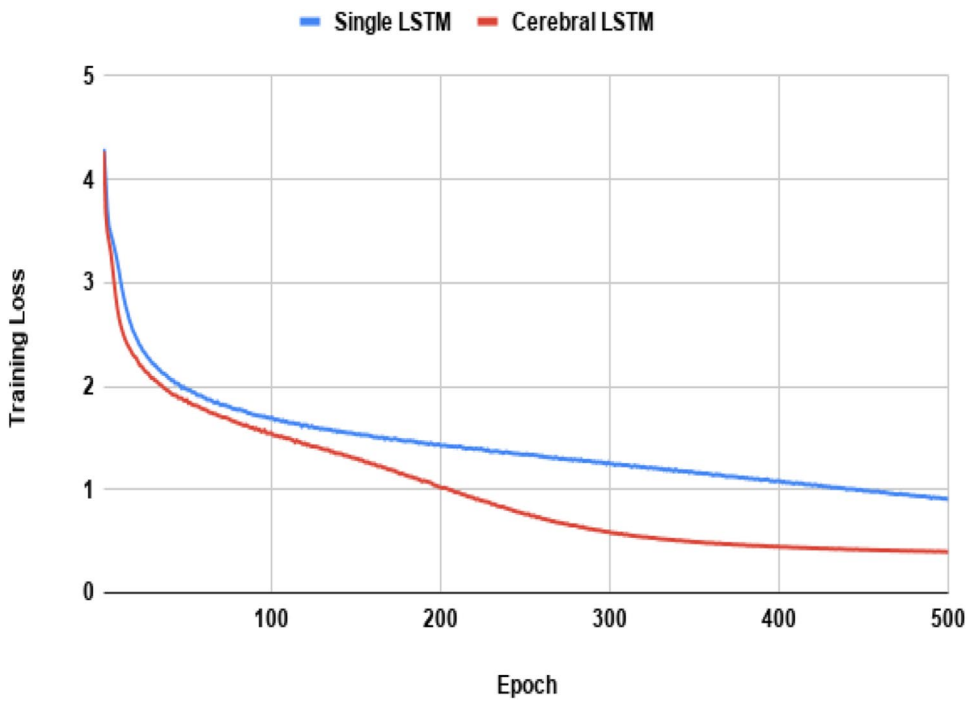

Fig. 13 Comparative study of training loss

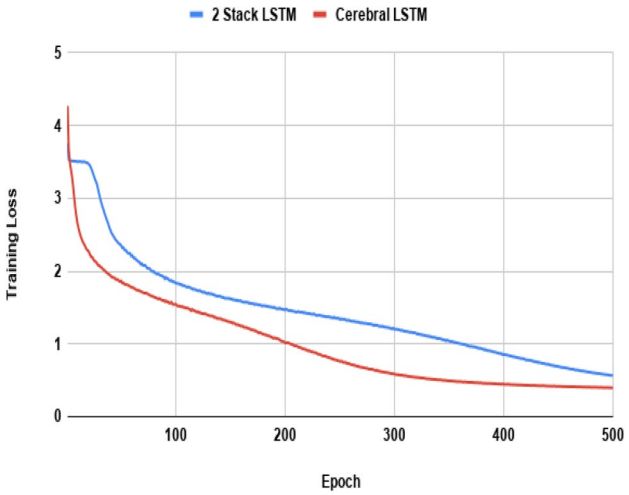

Fig. 14 Comparative study of training loss

We have provided all the experimental results in our GitHub repository [11] along with the dataset used to perform the comparative analysis.

Conclusion

Our proposed recurrent cell 'Cerebral LSTM' showed the ability to better understand data and has easily outperformed both single-LSTM- and two-stacked LSTM-based recurrent neural networks. Many variants of Cerebral LSTM can be designed using available varieties of LSTM cells including peephole LSTM. Further research work can be conducted on designing Cerebral LSTM-based stacked recurrent neural networks for designing deep learning architectures for understanding time-series data. Other recurrent cells including gated recurrent units can also be analyzed after modifying its internal connections similar to our cerebral structure.

Table 3 Sample of data generated by each recurrent unit

Single LSTM	Two-stacked LSTM	Cerebral LSTM
Lenny_Leonard: I'm sure serve big chief 'am there Lenny_Leonard: Ne need givan e	Buen… Thief_Wiggum: As a little room, it could pulce would yo, nive? Linder: Ho	Lenny_Leonard: What's the matter Home T? Moe_Szyslak: (MADDED) Jh…

References

1. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
2. Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int J Uncertain Fuzziness Knowl Based Syst. 1998;6:107–16. https://doi.org/10.1142/S0218488598000094.
3. Chung J, Gulcehre C, Cho KH, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Cornell University Library. 2014. http://arxiv.org/abs/1412.3555.
4. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 2005;18(5–6):602–10.
5. Cheng G, Peddinti V, Povey D, Manohar V, Khudanpur S, Yan Y. An exploration of dropout with LSTMs. In: Interspeech, Annual conference of the International Speech Communication Association, Stockholm; 2017.
6. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn. 2015;15:1929–58.
7. Huang A, Wu R. Deep learning for music. Cornell University Library. 2016. http://arxiv.org/abs/1606.04930.
8. Graves A, Mohamed A-R, Hinton G. Speech recognition with deep recurrent neural networks. In: Proceedings of international conference on acoustics, speech and signal processing. 2013. p. 6645–49.
9. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Proceedings of advances in neural information processing systems. 2014. p. 3104–3112.
10. Hossain Z, Sohel F, Shiratuddin F, Laga H. A comprehensive survey of deep learning for image captioning. Cornell University Library. 2016. http://arxiv.org/abs/1810.04020.
11. Github repository. https://github.com/mr-ravin/cerebral-rnn-experimental-results. Last accessed 31 Jan 2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.